DOCUMENT RESUME

ED 094 706 IR 000 901

AUTHOR Ogden, Philip M.
TITLE Monte Carlo Simulation of Counting Experiments.
PUB DATE 23 Oct 73
NOTE 8p.; Paper presented at the Rochester Area Colleges Workshop on the Use of Computers in Instruction (Rochester, New York, October 23, 1973)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Calculation; *Computer Programs; *Computer Science; Experiments; *Mathematical Models; *Probability Theory; *Simulation; Statistics
IDENTIFIERS *Monte Carlo Method

ABSTRACT

A computer program to perform a Monte Carlo simulation of counting experiments was written. The program was based on a mathematical derivation which started with counts in a time interval. The time interval was subdivided to form a binomial distribution with no two counts in the same subinterval. Then the number of subintervals was extended to infinity to form the Poisson distribution. Under restricted conditions the Gaussian (normal) distribution can approximate the Poisson. The computer program was written in the BASIC and FORTRAN languages and generates a pseudo-random number for each subinterval. These generated numbers are then compared with the Poisson and Gaussian distributions. A graphical display of the results is part of the program. A sample output is supplied in an appendix. (WH)

ED 094706
IR 000 901

MONTE CARLO SIMULATION OF COUNTING EXPERIMENTS

Philip M. Ogden
Division of Natural Science and Mathematics
Roberts Wesleyan College

BEST COPY AVAILABLE

I. INTRODUCTION

There are many phenomena that involve the counting of events under circumstances such that the probability of getting an event in a given time interval is independent of past history and is the same for any two time intervals of the same duration. Such conditions are often satisfied in nuclear or particle physics experiments, but are also satisfied in many other cases such as the counting of mutations in a large population or the counting of supernovae explosions in a galaxy. These conditions are quite well satisfied in nuclear radioactive decay if the activity is low enough so that depletion of the number of unstable nuclei is negligible. They are also generally satisfied in counting experiments using accelerators or cosmic rays if the beam and target conditions are constant.

In many counting experiments, it is not feasible to make a large number of repeated measurements in order to determine the statistical distribution of the results. This paper describes a computer program used to make Monte Carlo simulations of counting experiments in order to demonstrate the statistical distribution of such measurements. A BASIC language version for use on conversational systems and a FORTRAN language version for use on batch processing systems are available.

II. THEORY - THE POISSON DISTRIBUTION

When repeated observations of the type described in Section I are made, it is found that, in general, the results do not fit a Gaussian distribution. In-

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

stead, they fit the Poisson distribution described briefly in this section.[1,2]

Suppose that the total time interval T is divided into s subintervals with s large enough that the probability of getting more than one event in a given subinterval is negligible. One can eliminate the error introduced by neglecting multiple events if s is allowed to approach infinity. The probability of getting an event in a given subinterval can be written as

$$p = C(\frac{T}{s}) = \frac{k}{s} ,$$

where C and k are constants (for a given total time T). The probability of no event will therefore be

$$q = 1 - p = 1 - \frac{k}{s} .$$

The probability of getting one event in each of the first n subintervals and no events in the remaining (s-n) subintervals would be

$$p^n q^{(s-n)} = (\frac{k}{s})^n (1 - \frac{k}{s})^{s-n} .$$

However, there are many ways in which one can get n events, each with the probability above. There are s choices at first as to which subinterval will have an event, (s-1) choices for the next event, and so on until there are (s-n+1) choices for the nth event. Therefore, there are

$$s(s-1)(s-2) \ldots\ldots (s-n+1) = \frac{s!}{(s-n)!}$$

combinations of full and empty subintervals yielding a total of n events. Actually, we have counted each combination more than once since the n events are indistinguishable and there are n! ways of rearranging the n events without changing the condition of any subinterval. Therefore, we must divide by n! so that we obtain finally the probability of getting n events:

$$P(n, s, \frac{k}{s}) = \frac{s!}{n!(s-n)!} (\frac{k}{s})^n (1 - \frac{k}{s})^{s-n} .$$

At this point we have derived the binomial distribution which is just the probability of getting n events in s tries when the probability of success is (k/s). We wish now to take the limit of this binomial distribution as s approaches infinity. We will call this limiting form $P(n, k)$:

$$P(n, k) = \lim_{s \to \infty} \frac{s!}{n!(s-n)!} \left(\frac{k}{s}\right)^n \left(1 - \frac{k}{s}\right)^{s-n}$$

$$= \lim_{s \to \infty} \frac{s(s-1)\ldots(s-n+1)\, k^n \left(1 - \frac{k}{s}\right)^s}{n!\, s^n \left(1 - k/s\right)^n}$$

$$= \lim_{s \to \infty} \frac{\left(1 - \frac{1}{s}\right)\left(1 - \frac{2}{s}\right)\ldots\left(1 - \frac{n+1}{s}\right) k^n \left[\left(1 - \frac{k}{s}\right)^{-\frac{s}{k}}\right]^{-k}}{n! \left(1 - \frac{k}{s}\right)^n}$$

All the factors in parentheses approach one and the expression in the square brackets approaches the base of natural logarithms e. Therefore:

$$P(n, k) = \frac{k^n e^{-k}}{n!} .$$

This is the Poisson distribution. It is not difficult to show that[2]

$$P_{\text{total}} = \sum_{n=0}^{\infty} P(n, k) = 1 ,$$

$$\bar{n} = \sum_{n=0}^{\infty} nP(n, k) = k ,$$

$$\sigma^2 = \sum_{n=0}^{\infty} (n - \bar{n})^2 P(n, k) = k .$$

Therefore, we could write the Poisson distribution in the form:

$$P(n, \bar{n}) = \frac{\bar{n}^n e^{-\bar{n}}}{n!} .$$

Finally, we should point out that when $\bar{n}$ is large and n is not too far from $\bar{n}$, the Poisson distribution can be approximated quite well by a Gaussian distribution centered at $\bar{n}$ with standard deviation $\sqrt{\bar{n}}$. That is[3],

$$P(n, \bar{n}) \xrightarrow[\substack{\bar{n} \gg 1 \\ n \cong \bar{n}}]{} \frac{1}{\sqrt{2\pi\bar{n}}} e^{-\frac{(n-\bar{n})^2}{2\bar{n}}} .$$

III. COMPUTER PROGRAM

The computer program described here simulates phenomena of the type discussed in Section I by a Monte Carlo method using pseudo-random numbers generated by the computer. The total time interval is divided into N1 subintervals so that the probability of getting an event in a given subinterval is P1 = M/N1 where M is the predicted mean number of counts. The numbers M and N1 are entered as input by the user and should be chosen such that P1 is small, say .05 or less. This should be done in order that the error introduced by the neglect of multiple events in a subinterval will be small.

Ideally, the pseudo-random numbers will be distributed uniformly between 0 and 1 so that the probability that a given pseudo-random number lies between X and X + P1 will be just P1. For each subinterval a pseudo-random number is generated. If the number lies between 0.5 - P1/2 and 0.5 + P1/2, a count is recorded for that subinterval. These counts are accumulated to give a total count for that "experiment". This whole process is repeated N2 times where N2 is also entered as input by the user. After N2 "experiments", the results are accumulated in an array giving the number of "experiments" for each total count yield.

These Monte Carlo data are compared with the Poisson and Gaussian distributions. The mean of the Monte Carlo data is calculated and this quantity and the total number of "experiments" are used to determine a Poisson distribution for comparison. Similarly, the mean, standard deviation, and total number of "experiments" are used to determine a Gaussian distribution for comparison.

In both comparisons, the chi-square and the probability of exceeding this chi-square are calculated. The latter calculation is made using a method similar to that described by Bevington.[1]

The program has been written in a BASIC language version for use on conversational systems and a FORTRAN language version for use on batch processing systems. A sample output from the BASIC language version is listed in the Appendix. Program listings and more detailed descriptions are available from the author.

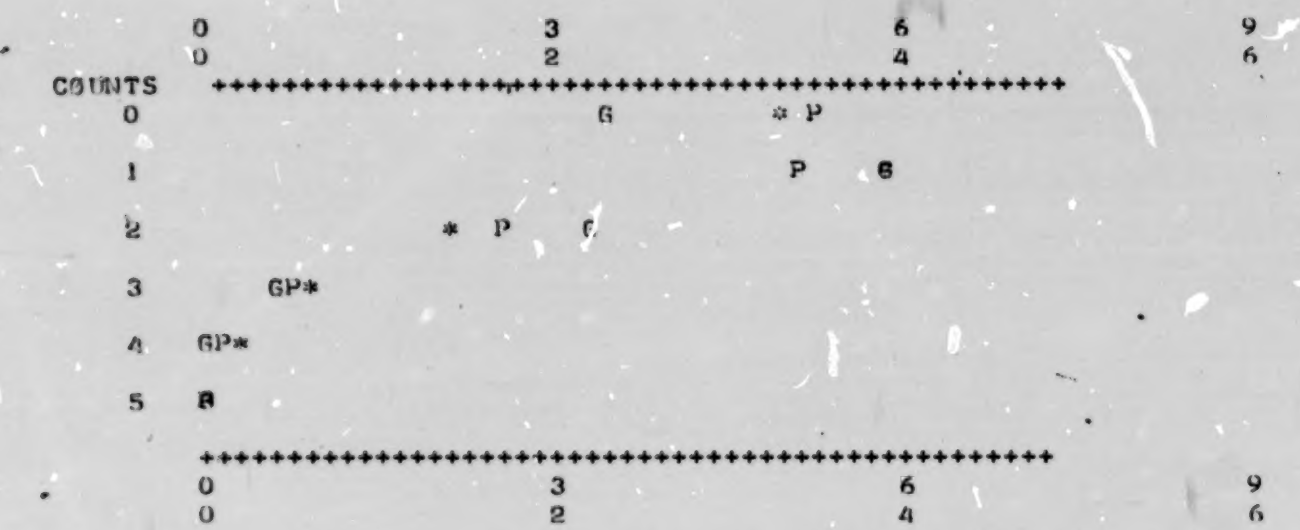
APPENDIX SAMPLE OUTPUT - BASIC LANGUAGE VERSION

```
WOULD YOU PLEASE ENTER TODAY'S DATE AS THREE NUMBERS SEPARATED
BY COMMAS? 8,5,71

MONTE CARLO SIMULATION OF COUNTING EXPERIMENTS

 8 / 5 / 71

THE PREDICTED MEAN CAN BE ANY NUMBER BETWEEN 0 AND 24.5 AND
NEED NOT BE AN INTEGER.  A VALUE OF 0 WILL TERMINATE THE
PROGRAM.  WHAT VALUE DO YOU WISH FOR THE PREDICTED MEAN? 1

ENTER A NUMBER BETWEEN 0 AND 1 AS A SEED FOR THE PSEUDO-
RANDOM NUMBER GENERATOR? .135793


IN THIS SIMULATION, THE TOTAL TIME INTERVAL IS DIVIDED INTO
SUBINTERVALS SO THAT THE PROBABILITY OF GETTING A COUNT IN A
GIVEN SUBINTERVAL IS THE PREDICTED MEAN DIVIDED BY THE
NUMBER OF SUBINTERVALS.  THE NUMBER OF SUBINTERVALS SHOULD
BE CHOSEN LARGE ENOUGH THAT THE PROBABILITY OF GETTING A
COUNT IN A GIVEN INTERVAL IS SMALL, SAY .05 OR LESS, IN
ORDER THAT THE PROBABILITY OF GETTING TWO OR MORE COUNTS IN
A SUBINTERVAL IS NEGLIGIBLE.  HOWEVER, TOO LARGE A NUMBER
OF SUBINTERVALS MAY CAUSE EXCESSIVE RUN TIME AND MAY ALSO
ACCENTUATE MINOR IMPERFECTIONS IN THE PSEUDO-RANDOM NUMBER
GENERATION.  THERE IS PROBABLY NO POINT IN CHOOSING THIS
NUMBER SO LARGE THAT THE PROBABILITY IS LESS THAT .005.
HOW MANY SUBINTERVALS DO YOU WISH? 100


IN CHOOSING THE NUMBER OF EXPERIMENTS OR MEASUREMENTS TO BE
SIMULATED, ONE SHOULD CHOOSE A NUMBER LARGE ENOUGH TO MAKE
THE RESULTS STATISTICALLY SIGNIFICANT, SAY 100 OR GREATER,
BUT NOT SO LARGE AS TO PRODUCE AN EXCESSIVE RUN TIME.  ONE
SHOULD KEEP IN MIND THAT THE RUN TIME IS PRIMARILY DETER-
MINED BY THE NUMBER OF RANDOM NUMBERS GENERATED AND THAT
THE NUMBER GENERATED IS THE PRODUCT OF THE NUMBER OF SUB-
INTERVALS AND THE NUMBER OF EXPERIMENTS.
HOW MANY EXPERIMENTS DO YOU WISH? 150

ENTER CODE FOR GRAPHICAL OUTPUT:
   0 - NO GRAPH
   1 - 1 GRAPH, 3 PLOTS(NOT POSSIBLE ON SOME SYSTEMS)
   3 - 3 GRAPHS
? 1
```

COUNTS	EXPERIMENTS	POISSON	GAUSSIAN
0	53	56.67323	37.28486
1	63	55.16194	62.16075
2	22	26.84548	35.19691
3	9	8.709867	6.762047
4	3	2.119401	.4409374
5	0	.4125767	9.758900E-3

PREDICTED MEAN 1
DATA MEAN .9733333
VARIANCE .9254586
STANDARD DEVIATION .9620076
EXPERIMENTS 150
EXPERIMENTS IN QUOTED RANGE 150
SUBINTERVALS 100
REDUCED CHI-SQUARE - POISSON FIT .683752
PROBABILITY OF EXCEEDING THIS CHI-SQUARE .6031027
REDUCED CHI-SQUARE - GAUSSIAN FIT 5.108731
PROBABILITY OF EXCEEDING THIS CHI-SQUARE 1.572132E-3

* - MONTE CARLO P - POISSON G - GAUSSIAN



MONTE CARLO SIMULATION OF COUNTING EXPERIMENTS

6 / 5 / 71

THE PREDICTED MEAN CAN BE ANY NUMBER BETWEEN 0 AND 24.5 AND NEED NOT BE AN INTEGER. A VALUE OF 0 WILL TERMINATE THE PROGRAM. WHAT VALUE DO YOU WISH FOR THE PREDICTED MEAN? 10

ENTER A NUMBER BETWEEN 0 AND 1 AS A SEED FOR THE PSEUDO-RANDOM NUMBER GENERATOR? .6

REFERENCES

1. Bevington, Philip R., Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York (1969), pp. 36-42, 189-193.

2. Wilson, E. Bright, Jr., An Introduction to Scientific Research, McGraw-Hill, New York (1952), pp. 191-195.

3. Beers, Yardley, Introduction to the Theory of Error, Second Edition, Addison-Wesley, Reading, Massachusetts (1957), pp. 46-49.